Eukaryotic Transcription Regulation

Introduction

Eukaryotic gene expression is actually quite more complex than we let on in the previous lesson. We know there's a promoter region, a TATA box, and exons and introns. Based on processing, we can get different protein products from the same stretch of DNA. In this lesson we're going to talk more about how genes are regulated on a larger scale and how the nucleus "knows" where to put RNA polymerase, and cover some of the steps that regulate gene expression. Regulation is all about the binding of proteins. More binding, more expression. Binding requires a euchromatic state. We start with some global changes in overall DNA structure (acetylation of histones, methylation of cytosines) and then move into the obnoxiously complex promoters and enhancers.

Chromatin 1: Histones and Acetylation

Euchromatin is open, accessible DNA that can be transcribed. Heterochromatin is closed, tightly packed, inaccessible DNA that cannot be transcribed. The first step to get to translation is to open the DNA to euchromatin.

Histone acetyltransferases (HATs) transfer an **acetyl group** to **lysine** on the histone. This physical barrier (the acetyl group) removes the positive charge and causes both electrostatic inhibition and physical disruption to the dense packing of that DNA into the 30nm nucleosome.

Histone deacetylase (HDAC) undoes this.

There is a constant duel between HATs and HDACs that keeps the accessibility of DNA in constant flux.

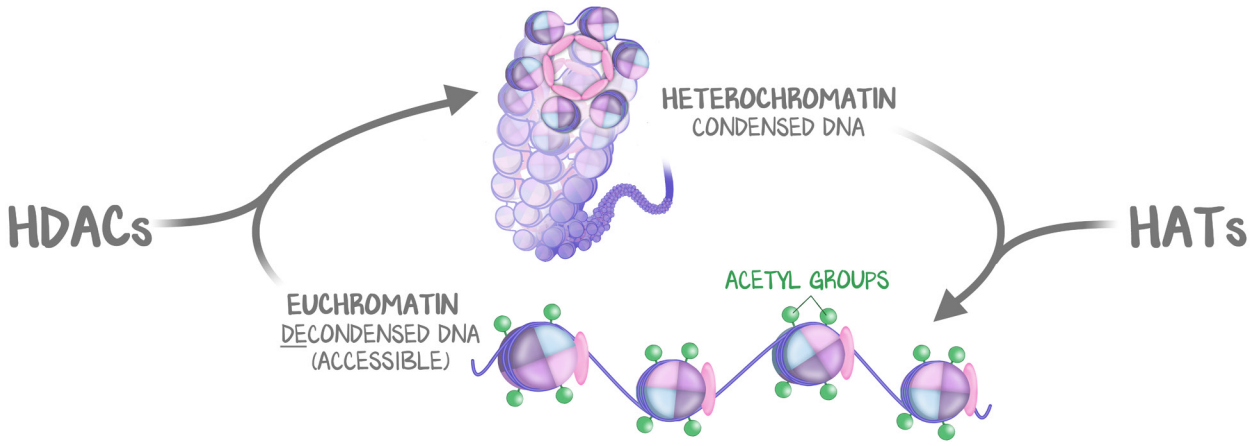

Figure 10.1: HATs and HDACs
Shows the battle between HDACs, which de-acetylate histones allowing them to condense, and HATs, which acetylate histones preventing them from condensing, forcing them open to euchromatin.

Chromatin 2: Methylation

Cytosine residues in DNA can be methylated to produce 5-methylcytosine. Methylated cytosines influence the **GC-rich domains** in the proximal promoter. When methylated, **genes are transcribed less**. There is a big to-do about methylation, in that methylation is plastic—the state of mom when she releases her zygote can actually influence the genetic coding during pregnancy; a sort of "inheritance" of more than just genes, but gene expression.

This is brought up on the exam by the difference between **Prader-Willi** and **Angelman** syndromes. Both syndromes result from deletions of the same region of **chromosome 15**. Which one, depends on the **methylation state** of the sister chromosome (the one that isn't deleted). Males methylate a different set of genes than females. If the **deletion is of dad's** chr 15, **mom's methylated 15** remains, resulting in **Prader-Willi**. If the **deletion is of mom's** chr 15, **dad's methylated 15** remains, resulting in **Angelman syndrome**.

A Preview of Initiation: Basal Transcription Apparatus

The **basal transcription apparatus (BTA)** is the combination of the common general transcription factors and RNA polymerase. As the DNA becomes more euchromatic at this exact location, as it literally unrolls, the unrolling reveals the **promoter**. The preinitiation complex begins to assemble. Starting with the core promoter with its **TATA box**, the TATA-binding protein domain of transcription factor IID binds to TATA. More general transcription factors bind the upstream promoter-proximal elements. Together, promoter, TBP, TFIID, and other transcription factors local to the promoter form the **preinitiation complex** (PIC). The PIC acts as the beacon for the site of attachment for **RNA polymerase**. Addition of the polymerase to the preinitiation complex makes it the **basal transcription apparatus**. Once attached and assembled, RNA polymerase starts to transcribe, but slowly.

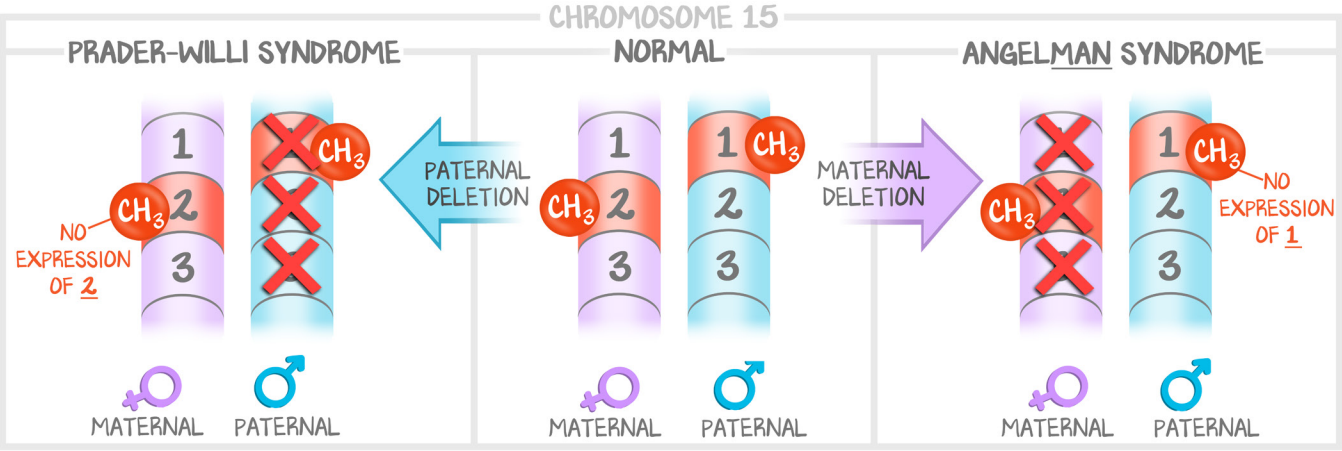

Figure 10.2: Methylation and Imprinting
If the maternal gene is deleted (maternal 1,2,3) the methylation state of the paternal gene (1-methyl) means that NO expression of section 1 occurs. Mom's copy got deleted, so doesn't exist, and dad's is methylated so it isn't expressed. So only 2,3 are coded. If the paternal gene is deleted (paternal 1,2,3) the methylation state of the maternal gene (2-methyl) means that NO expression of gene 2 occurs. Dad's copy doesn't have 2 at all, and mom has it methylated. Only 1,3 are coded.

If all that happens is the basal transcription apparatus forms, the gene is said to be **constitutively on**—always on and with a meager basal rate. More cell-specific, more powerful gene regulation comes from cell-specific transcription factors and distal DNA sequences—either downstream or upstream of the promoter. These are enhancers.

Regulation by General Transcription Factors: The Promoter

RNA polymerase must recognize the appropriate point at which to start transcription and the strand of the DNA to transcribe (the template strand). RNA polymerase also must recognize which genes to transcribe, because transcribed genes are only a small fraction of the total DNA. RNA polymerase knows where and what to bind because of the promoters, consensus sequences, and transcription factors that make the **preinitiation complex**.

Promoters are sequences in DNA that determine the start point and the frequency of transcription. They require **transcription factors** to bind the **DNA sequence** to form the preinitiation complex.

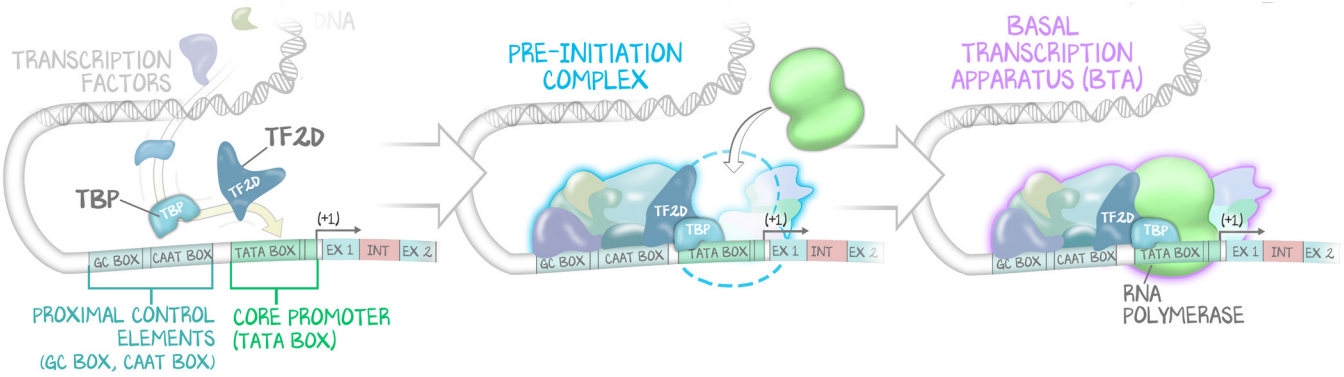

Figure 10.3: PIC and BTA
With all the elements in place except RNA polymerase—enhancer sequences binding activators connected to the promoter region by coactivators, TATA box bound by TBP, etc.—the preinitiation complex is ready for gene transcription to occur. Once RNA polymerase binds, transcription can start, and the basal transcription apparatus is ready.

The promoter is a combination of the **core promoter**—the **TATA box** (-25)—and the **proximal control elements**, also called consensus sequences. These upstream proximal control elements are the **CAAT box** (-75) and a **GC-rich sequence** (about -150). These promoter regions have transcription factors that bind to these sequences and are part of the basal transcription unit. These regions of regulation in the proximal promoter region are said to be **general**, as are the **transcription factors** deemed **general**. "General" means *"the transcription factors and regions they bind are common to all cells, regardless of their state."* The idea is that we have to manage the **same structures** to get **RNA polymerase started**. They are both necessary and sufficient for a basal rate. The big to-do is in the enhancer regions, which can **drastically increase** or **decrease transcription**.

The TATA box is the principal event that must occur to even start the process of initiation. The **TATA box is highly conserved**. It is found at -25 base pairs in eukaryotes. Any mutation of the TATA box results in the total inability of transcription. When the **TATA-binding protein** (TBP) subunit of **transcription factor II D** (TFIID) binds to the **TATA box** it induces a **bending of DNA**. This shape change allows transcription factors access to the DNA. Other named-and-lettered transcription factors add to the construct to form the basal transcription complex, but memorizing that level of detail doesn't seem to be worth it.

Regulation by Cell-Specific Transcription Factors: Enhancers

Enhancers are DNA sequences that are quite distal to the proximal promoter. Enhancers are the DNA sequences. **Activators** are proteins that bind to enhancers and also to RNA polymerase. Activators give the go signal for RNA polymerase. There are technically also silencer regions that bind repressors, but are so rare, we leave them out. Just know that an activator gives the go signal and repressors inhibit that signal.

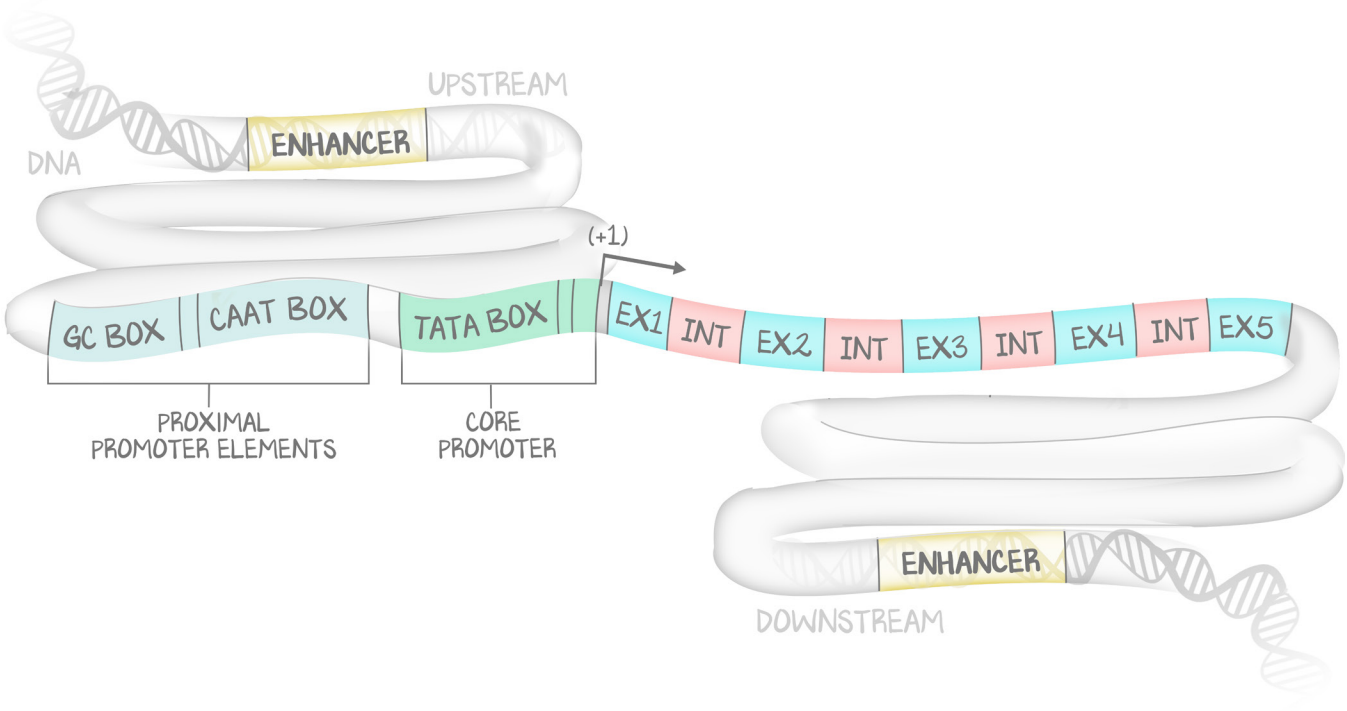

Figure 10.4: Promoter and Control Elements
The proximal control elements can respond to general or specific transcription factors. The core promoter is general—every cell has the same basic start point. The enhancers, discussed next, are distal to the promoter and are almost always specific to the cell. The proximal control elements are a mix of general and specific—the farther from transcription start, the more likely it will be mediated by the binding of cell-specific transcription factors.

Enhancers are thousands of base pairs away from the gene. They can be either **upstream** or **downstream** from the gene. Enhancers interact with the proximal promoter region by **bending the DNA** to come into contact with the local area. The signal from enhancer-regions-activated-by-activators is communicated to the basal transcription apparatus through **coactivators**, intermediaries.

Most importantly, **enhancers** act in a **tissue-specific manner** and, as **DNA-binding sites**, are capable of responding to alterations in the cell's environment. Enhancers are effectively **receptors** for **gene-specific transcription factors**, the level of which varies based on extracellular signaling and cytoplasmic levels of phosphorylation. Because these proteins are themselves transcribed, the function and level of activators can vary cell to cell, cycle to cycle, and therefore gene regulation at the cell-type level can be modulated by enhancers.

Where the proximal promoter region has common, general transcription factors binding to nearly ubiquitous promoter elements common to all genes, the enhancer regions allow for **cell-specific proteins** to bind **cell-specific enhancer regions**.

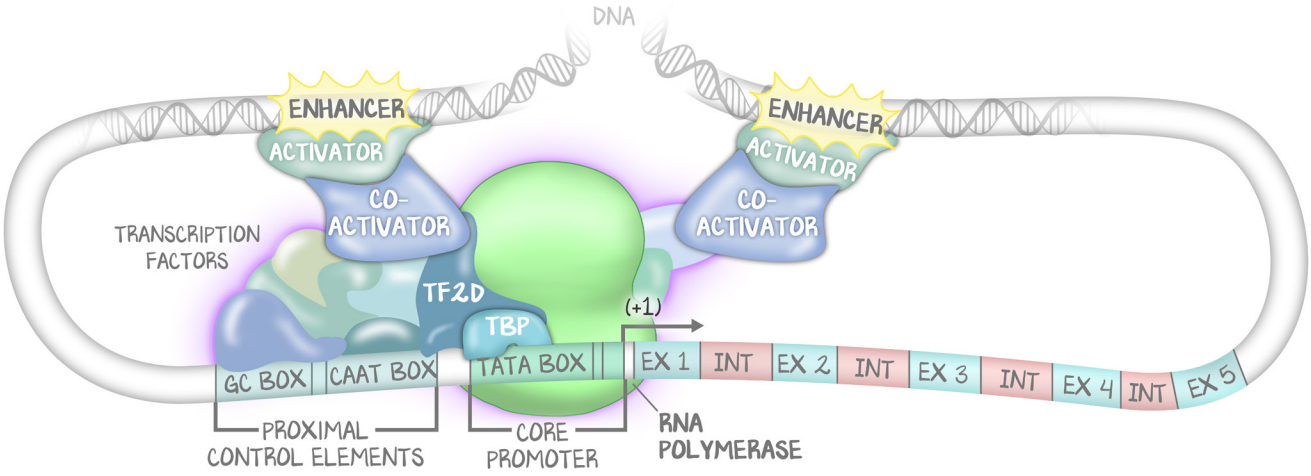

Figure 10.5: Enhancers
Enhancer sequences are in DNA to recruit activators (from cytoplasmic activation or steroid receptors), which in turn act through coactivators to stimulate the proximal promoter. These are often cell-specific transcription factors. The enhancers are so distal to the gene being transcribed that the DNA actually loops back around on itself in order to interact with the promoter.

Transcription Factors

Transcription factors (TFs) are the **activators** discussed above. "Transcription factor" therefore is a generalized term, and means the same thing as "protein" or "gene regulator." Transcription factors **bind DNA** at a specific **binding site** called a **response element** (these were the CAAT box, and GC-rich and enhancer regions). Transcription factors are **modular** in that they can be **activated** (readily binds to its response element) and inactivated (turned off) based on the cell's current environment and demands. This modularity is often in response to cellular events and will help direct transcription based on the cell's needs, type, and current resources.

TFs can either **recruit chromatin** (HATs) or **interact with RNA polymerase II**, stabilizing the preinitiation complex. It's the latter we will mostly discuss.

TFs that bind the **proximal promoter** are usually **general** and **common to all genes**. TFs that bind to **enhancers** are usually **cell-specific**. Everything that is a protein, that binds to DNA, and influences the transcription of DNA is a TF. TFs are the ligands for their receptor, the response element, found in the major grooves of helical DNA.

There are three commonly tested DNA-binding domains.

The **helix-turn-helix** refers to **homeodomains** or **Pax** proteins, used to differentiate cells in utero. DO NOT LEARN DETAILS. Simply put, if a child has dysfunction in many organs with developmental delay, and the question asks which gene-expression method, choose homeobox (HOX) or paired-box (PAX) genes.

The **leucine zipper** is a motif characterized by a leucine every seventh position. Leucine zippers dimerize and "grip" the DNA through basic amino acids (arginine and lysine), which easily bind the negative backbone of DNA. An example is glucagon's effect on hepatic gluconeogenesis. Glucagon binds to membrane receptors, then translates that signal intracellularly. Glucagon is carried forward by a cAMP-dependent activator protein. A rise in cAMP activates protein kinase A, which **phosphorylates**, thereby activating **CREB** (the leucine zipper). It binds to its cAMP response element (via the leucine zipper), which activates gene activity. Glucagon acts through the CREB to increase gluconeogenesis.

The **zinc fingers** are found in steroid hormone TFs as well as the steroid hormone response element. The **steroid hormone TF** (lipophilic ligand) binds to the **hormone response element** (the DNA sequence, the "receptor"). This binding acts as an activator at an enhancer region. In the liver, cortisol induces gluconeogenesis. Cortisol, a steroid hormone, can pass through lipid bilayers. It arrives at the cell, gets to the nucleus, and binds through zinc fingers to the glucocorticoid response element, which acts as an enhancer to the PEPCK gene. The PEPCK gene activates, promoting gluconeogenesis.

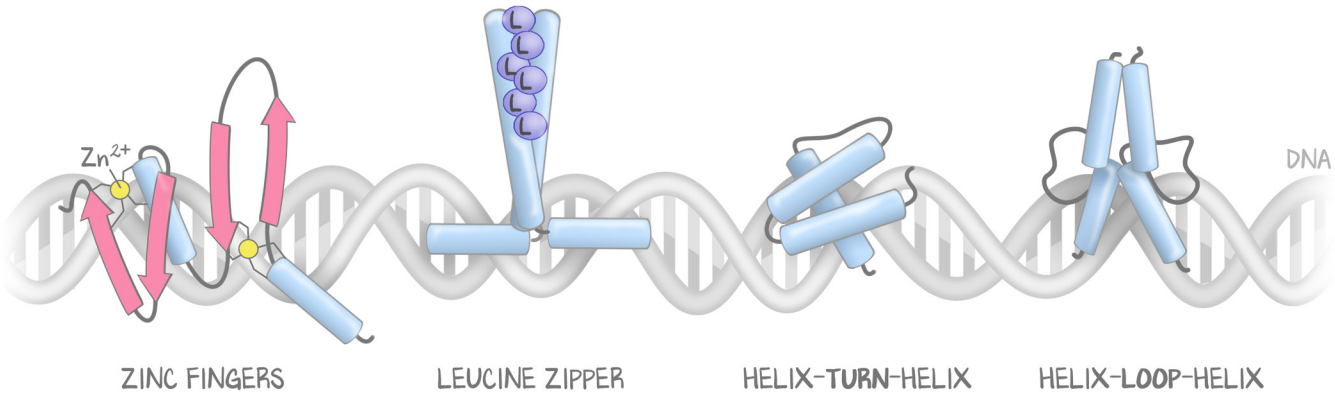

Figure 10.6: Transcription Factors
Illustration of the different types of Transcription factors shapes and their interaction with DNA.

Chicken and the Egg

A transcription factor is a transcribed protein with its own promoter and enhancers. Transcription factors alter the expression of other genes. There are so many mechanisms to activate or inhibit an activator, and worse, it isn't "on" and "off." It's the relative expression of transcription factors that causes a relative increase or decrease in this gene, and that same transcription factor might function differently on another gene. Don't try to figure out what I just wrote. Just realize that the interaction of hundreds of transcription factors is taking action on thousands of promoter regions all at the same time. You cannot keep them straight. Don't try. Focus on THIS transcription, and think of it as sequential events. You're more likely to keep from getting confused.